

Deep Convolutional Spiking Neural Network Based Hand Gesture Recognition

Weijie Ke

*Electronic & Electrical Engineering
University of Strathclyde
Glasgow
weijie.ke@strath.ac.uk*

Yannan Xing

*Electronic & Electrical Engineering
University of Strathclyde
Glasgow
yannan.xing@strath.ac.uk*

Gaetano Di Caterina

*Electronic & Electrical Engineering
University of Strathclyde
Glasgow
gaetano.di-caterina@strath.ac.uk*

Lykourgos Petropoulakis

*Electronic & Electrical Engineering
University of Strathclyde
Glasgow
l.petropoulakis@strath.ac.uk*

John Soraghan

*Electronic & Electrical Engineering
University of Strathclyde
Glasgow
j.soraghan@strath.ac.uk*

Abstract— Novel technologies for EMG (Electromyogram) based hand gesture recognition have been investigated for many industrial applications. In this paper, a novel approach which is based on a specific designed spiking convolution neural network which is fed by a novel EMG signal energy density map is presented. The experimental results indicate that the new approach not only rapidly decreases the required processing time but also increases the average recognition accuracy to 98.76% based on the Strathclyde dataset and to 98.21% based on the CapgMyo open source dataset. A relative comparison of experimental results between the proposed novel EMG based hand gesture recognition methodology and other similar approaches indicates the superior effectiveness of the new design.

Keywords— *convolutional spiking neural network, gesture recognition, EMG, signal processing*

I. INTRODUCTION

Hand gesture recognition which aims at assisting hand/partial hand amputees represents a challenging task in the biomedical and artificial intelligence fields[1]. In past years, several methodologies have been applied to address such problem. The machine learning based techniques are most commonly used for this task. Luan et al. [2] presented a LRCN deep neural network which is a combination of LSTM and CNN to address the dynamic hand gesture recognition problem for the robot arm control. Zhang et al. [3] successfully deployed conventional feature extraction-classification method to EMG signal based gesture dataset. In their work, many statistics feature is considered such as MAV, RMS, SSC, WL. Another neural network based gesture recognition work is delivered in [4], which employed 3DCNN and LSTM with FSM to constructed a context aware classification model.

Methodologies based on conventional electromyogram signal processing techniques such as wavelet transform processing produce very good results, which allow recognition systems using pre-trained features (information) instead of conjunctures when making judgements[2]. Despite such good accomplishments, there are still two main issues which exist in the gesture recognition field. One of them is the limitation in further improving recognition accuracy as it is difficult to extract features from multi-sensory EMG signals

[3, 4]. The second is the delayed reaction time due to complex signal pre-processing [5].

In this paper, a novel approach which solves these issues by combing EMG selecting methodology with convolutional spiking neural network will be presented. Our approach shows a significant improvement in processing speed and hand gesture recognition accuracy compared with previous hand gesture recognition methodologies.

The remainder of the paper is organised as follows: Section 2 provides some background on the EMG signals and convolutional spiking network used. Section 3 indicates the methodology for creating the common energy-density map and the structure of the applied neural network. Section 4 includes details about the experimental setup for applied datasets and the discussion of the experiment results of our methodology with comparisons with other approaches. Section 6 contains the conclusion.

II. BACKGROUND

A. sEMG Signals

Electromyography (EMG) involves the measurement of electrical activity from functional skeletal muscles. The arm movements usually are triggered by the central nervous system, which is controlled by the brain sending signals to the forearm. The points where muscle fibers are triggered by the nerves are called motor units (MU), and these transmit electrical signals which cause muscles to contract and relax. The voltage generated by the motor units is called motor unit action potential (MUAP)[6].

Currently, the EMG signals are used in medical examinations for evaluating the health condition of muscles and the nerve cells that control them. Diseases such as disorders affecting the connection between nerves and muscles are easily identified through EMG analysis. Furthermore, EMG signals are also commonly utilized as control signals for prosthetic devices (prosthetic hands, arms, and lower limbs) [7].

Nowadays, surface and intramuscular are the two types of sensors frequently used for EMG signal recording. Surface sensors, also known as sEMG sensors, record muscle activity from above the skin, providing data from the motor units that are near the skin. Normally, one sEMG sensor comprise two

electrodes. One of the electrodes is used as a reference and the other one is used to measure the potential difference with respect to the first electrode. Other configurations, such as a bipolar sensor, also exist. Intramuscular EMG sensor is generally performed by inserting a fine wire directly into a muscle, measuring the activity of a single or various motor units. In this configuration a surface EMG sensor is used as a reference point [8].

B. Convolutional Spiking Neural Network

A convolutional spiking neural network usually contains two main parts: spiking convolution operation and spiking encoding. The spiking encoding performs the encoding of pixel intensity value of the input images to spike train events and the spiking convolution operation extra and learn the features through training. The rule of a convolution operation is shown in equation 1:

$$y[m, n] = \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} x[i, j] \cdot h[m - i, n - j] \quad (1)$$

where m represents the horizontal coordinate of the picking point and n is the vertical coordinate of the picking point; i and j indicate horizontal and vertical coordinates for all pixels inside the input image [9]. The key of the convolutional operation is the 2-D kernel (filter) size and stride, i.e. kernel moving steps when computing the convolution.

The spiking encoding can be achieved through two approaches: rate coding and latency coding. Rate coding indicates that the information is transferred to the spike train that the frequency is proportional to the value of original data content: the higher frequency of the spike trains the more information is carried. Usually the mean rate of a Poisson process is applied to determine the spike rate in this case. Latency coding refers to the information is encoded by the spikes' arriving time: earlier arrived spikes contains more information compared with later ones[10].

In this paper, a neuron model named Leaky integrate and fire (LIF) is applied to achieve spiking encoding task. A LIF model is formed as a parallel combination of a 'leaky' resistor (conductance, g_l) and a capacitor (C). The mathematical representation of a LIF model is shown in equation 2:

$$C \frac{dv}{dt} = -g_l(V(t) - E_l) + I(t) \quad (2)$$

When a current source $I(t)$ applied, the capacitor will be charged and a potential $V(t)$ will be generated. The generated $V(t)$ will be compared with a pre-settled threshold V_{th} . Once the potential exceeds the threshold, a spike will be issued[11]. Meanwhile the capacitor discharges to a resting potential E_l to avoid misfiring of the spikes of the same information[12].

C. Common Energy-Density-map

In our previous work [11] it was shown that not all 128 sensors are useful when generating energy density maps, as some of the information, such as outliers and signals not directly generated by the corresponding muscle action, would misleading the neural network. A methodology was created to

reduce their impact. The same action is also taken in this work. Initially the optimal applied sensor numbers for each gesture is determined. The procedure for selecting these optimal numbers is as follows. Begin with the sensor with the highest signal (brightest color), a partial energy-density map is formed. The second partial energy-density map uses the two highest signal sensors (this is including the first sensor). This process of selecting partial energy-density maps continues, each time including the sensor with the next highest signal. The process stops when 50 (out of 64) sensors are selected for each band (100 sensors in total) – the last 14 sensors with very low signal strengths are ignored[6].

This current work focusses on locating the common sensors (sensors from common muscle areas for each gesture for all subjects) to evaluate common features. The common sensor locating methodology shown in algorithm 1 was used on these partial energy-density maps so as to determine the common active sensors for each gesture and, thus, evaluate the optimal number of sensors which would provide the best results.

Algorithm 1: Common Sensor Locating Algorithm

Input: partial energy-density maps for all subjects' gestures for all numbers of sensors considered

Parameter: image type: jpeg; image size: 8×8

Output: common energy-density maps

- 1: collect all partial energy-density maps for same sensor number for all subjects' gestures
 - 2: select one partial energy-density map for one gesture from one subject
 - 3: intersect the energy-density map of the selected subject's gesture from all remaining energy-density maps (from all other users) for the same gesture to generate common energy-density maps between the chosen subject and each of the other subjects
 - 4: repeat steps 2 to 3 for all other subjects
 - 5: repeat steps 2 to 4 for all remaining gestures
 - 6: repeat steps 1 to 5 for all remaining sensor number selections.
-

Using this Common Sensor Locating Algorithm, it was possible to determine, and use for classification, the readings from sensors which were common to all users in each case. Moreover, the algorithm obtains the common energy-density features through selecting common active sensors for all users for each gesture - 50 sensors were applied for the experiment for both datasets and the readings from the best 35 readings from these sensors were selected in this study.

TABLE I. Common Active Sensors for Strathclyde Dataset

Gestures	Lower Band Common Sensor Locations	Upper Band Common Sensor Locations
Palm up	48,49,50,53,54,55	103,104,105,117,118,119,125,126,127,128
Palm down	24, 27, 50	107, 108, 109, 117,118,119,121,122,123,124,125,126,127,128
Tripod closed	3, 4, 5, 6, 54, 55	95,96,97,105,106,107,108,117,118,122,123,126,127,128
Lateral	1, 2, 3, 4, 5, 6	67,105,106,113,114,117,118,120,121,122,123,124,126,127

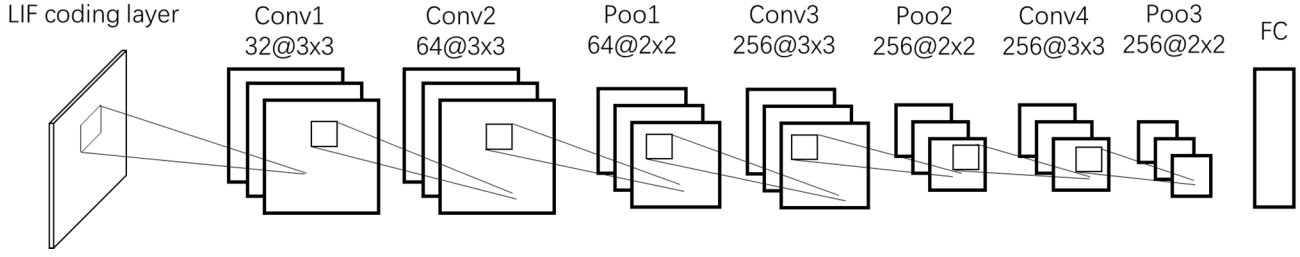


Fig.1. The Architecture of proposed network

Point	3, 4, 5, 6, 20, 21, 22	117,118,120,121,122,123,124,126,127,128
Flexion	2, 3, 4, 20, 21, 22	91,92,93,94,95,96,97,117,118,126,127,128
Extension	46,47,48,49,53,54,55,56,57	95,96,97,111,112,113,117,118,119,125,126,127
Close	3,4,5,6,7,8,27,28,29,49,50	110,111,112,113,117,118,119,125,126,127

Table I indicates the common active sensors for the Strathclyde dataset after applied algorithm (1. palm up; 2. palm down; 3. tripod closed; 4. lateral; 5. point; 6. flexion; 7. extension; 8. close).

Table II indicates the common active sensors for the CapgMyo dataset after applied algorithm (1. thumb up; 2. extension flexion; 3. flexion extension; 4. thumb opposing; 5. abduction; 6. fingers flexed; 7. pointing index; 8. adduction).

TABLE II. Common Active Sensors for CapgMyo Dataset

Gestures	Lower Band Common Sensor Locations	Upper Band Common Sensor Locations
Thumb up	2, 3, 4, 5	113,114,117,118,120
Extension flexion	2, 3, 5, 46,47,48,49,53	95,96,97,111,112,113,117,118,119
Flexion extension	2,3,4,5,46,47,48,49,53,54,55,56	95,96,97,111,112,113,117,118,119,125,126,
Thumb opposing	24,25,27,47,48,49,50	117,118,119,121,122,123
Abduction	2,3,4,5,46,47,48,53,54,55,56	95,96,97,111,112,113,117,118,119,125,126,127
Fingers flexed	3,4,5,6,7,8,27,28,29,49,50	110,111,112,113,117,118,119,125,126,127
Pointing index	3,4,5,20,21,23	117,118,120,121,122,123,124,126,127,128
Adduction	48,49,50,53,54,55	103,104,105,117,118,119,125,126,127,128

For each gesture in the Strathclyde dataset [11], 340 partial energy-density maps (30 partial energy-density maps from 9 able handed subjects and 5 from 14 amputees) were obtained for each set of sensors selected (from 1 to 50). In total, 136000 partial energy-density maps are obtained (340 per-gesture \times 8 gestures \times 50 sensors selection conditions). After applying the common sensor locating algorithm for each set of sensors selected, 57630 common energy-density maps were obtained per-gesture. For all, 23,052,000 common energy-density maps were applied (57,630 per-gesture \times 8 gestures \times 50 sensors selection conditions).

As for each gesture in the CapgMyo dataset, 38 partial energy-density maps were obtained for all subjects for each set of sensors selected (from 1 to 50). With 8 gestures performed by each subject, 15200 partial energy-density maps are obtained (38 per-gesture \times 8 gestures \times 50 sensors selection conditions). With applied common sensor locating algorithm, 296,400 common energy-density maps were applied (741 per-gesture \times 8 gestures \times 50 sensors selection conditions) in total.

Examples of acquired common energy-density maps are shown in the left side of figures 4(a) to 4(f). The white colours represent the location of selected common sensors which contain the common features. The brighter the colour is, the higher the energy density that sensor contains. The black colours stand for the location of the sensors that removed after common energy-density map generation procedure i.e. uncommon sensors.

D. Applied Convolutional Spiking Neural Network

After the common energy-density maps generation processing procedure, the pre-designed convolutional spiking neural network is applied on obtained common energy-density maps for each gesture. The applied convolutional spiking neural network shares a similar construction with normal convolutional neural networks except input and output layers. Figure 1 indicates the network structure. The first layer (input layer) is formed by a group of leaky integrate-and-fire (LIF) neurons which converts each pixel of the input image into spikes (spike images). The second layer is a convolutional layer with 32 kernels (size 3×3). The third layer is another convolutional layer with 64 kernels (size 3×3). The fourth layer is an average pooling layer (size 2, strides 2). The fifth layer is also a convolutional layer with 128 kernels (size 3×3). The sixth layer is another average pooling layer which has the same parameters as previous one. The seventh layer is the last convolutional layer with 256 kernels (size 3×3). The eighth is the last pooling layer (size 2, strides 2). The last part of the network is a dense layer (fully connected layer). Moreover, each convolutional layer is followed by an empty layer which only transfers the output of the previous layer to the next layer without any process.

III. EXPERIMENT RESULTS

Two datasets with different hand gestures are applied in this research. One is the self-collected Strathclyde dataset and the other one is the CapgMyo dataset. Both datasets are obtained through similar equipment and methodology.

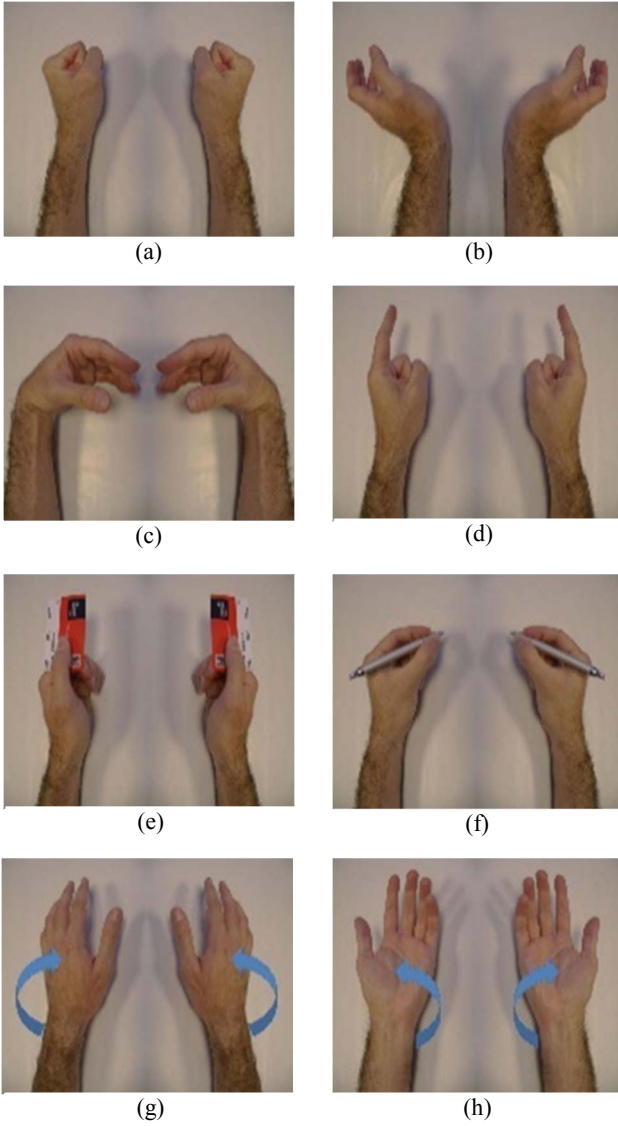


Fig. 2. Classified gestures (a) close; (b) extension; (c) flexion; (d) point; (e) lateral; (f) tripod closed; (g) palm down; (h) palm up.

A. Strathclyde Dataset

We firstly evaluated the proposed network with the self-collected hand gesture dataset (Strathclyde dataset).

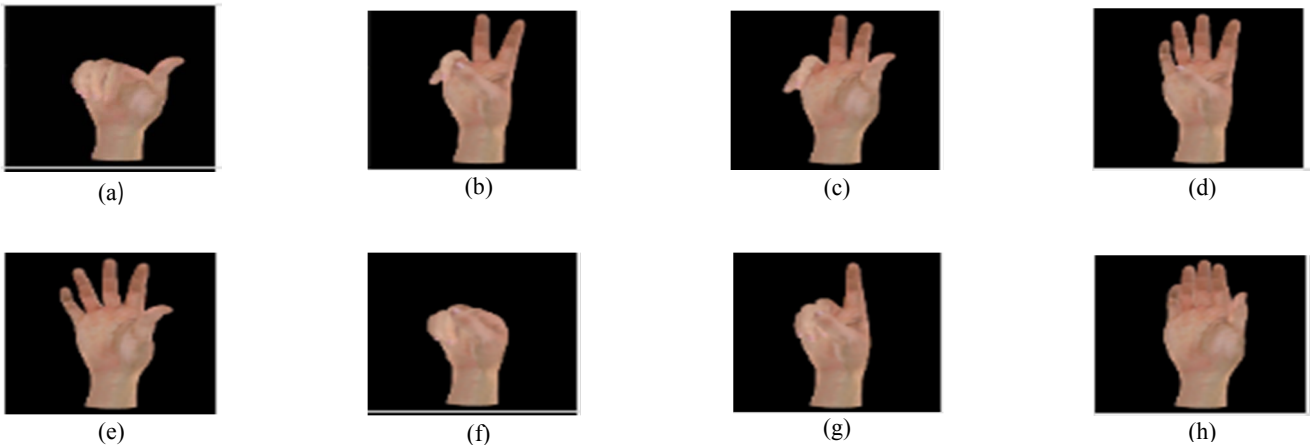


Fig. 3. (a) Thumb up; (b) Extension of index and middle, flexion of the others; (c) Flexion of ring and little finger, extension of the others; (d) Thumb opposing base of little finger; (e) Abduction of all fingers; (f) Fingers flexed together in fist; (g) Pointing index; (h) Adduction of extended fingers.

This dataset was obtained using sEMG sensors. The sEMG signal consists of the summation of the different MUAP within its reach plus some additional external noise. The Strathclyde dataset was recorded using two HD EMG arrays, each comprising 64 sensors. One EMG array was placed on the surface of extensor forearm muscles, while the other was placed on the flexor muscles. Of the 22 participants involved in the study 9 were able bodied and the rest were amputees with different levels of hand/lower arm amputations [5].

Sensor locations remained the same for both amputees and able handed participants when performing the same test.

There were 12 distinct movements performed by all subjects, which were all formed from rest and maintained for 3 to 5 seconds. Each gesture was repeated 10 times for able-bodied subjects and 5 times for amputees in a random order. However, some of the gestures performed by the amputees were different compared to the able bodies. For example, some of the amputees were unable to perform a grip. Hence, only the 8 hand activities, which both sets of subjects could achieve, were used for analysis and classification. All 8 gestures shown in Figure 2 were performed under the same conditions[6].

The recorded signals were amplified and sampled at 2048 Hz. After collection, the sampled data is subsequently passed through built-in hardware filters comprising a high-pass filter with cut off frequency 3Hz, and a low-pass filter with cut off frequency 900 Hz as significant EMG activity happens between 5 Hz and 450 Hz. A 3rd order Butterworth digital filter was employed to remove electronic equipment noise and motion artefacts. Signals are modified by Hamming windows with length 250ms, 50 percent overlap for the last processing step[8]. The examples of the testing results are shown in Figure 4(a), 4(b) and 4(c).

Among the figures, each number represents a pre-defined hand gesture/movement (0: palm up; 2: palm down; 3: tripod closed; 4: lateral; 5: point; 6: flexion; 7: extension; 8: close) with a specified colour. The vertical axis of the left part of each image indicates the possibility of which gesture is the input common energy-density map represents and the horizontal axis indicates the processing time in seconds.

Among all figures, the possibility of each input gesture always starts at 0 in the beginning. With an average 0.002s processing time, the lines which represents the gestures with

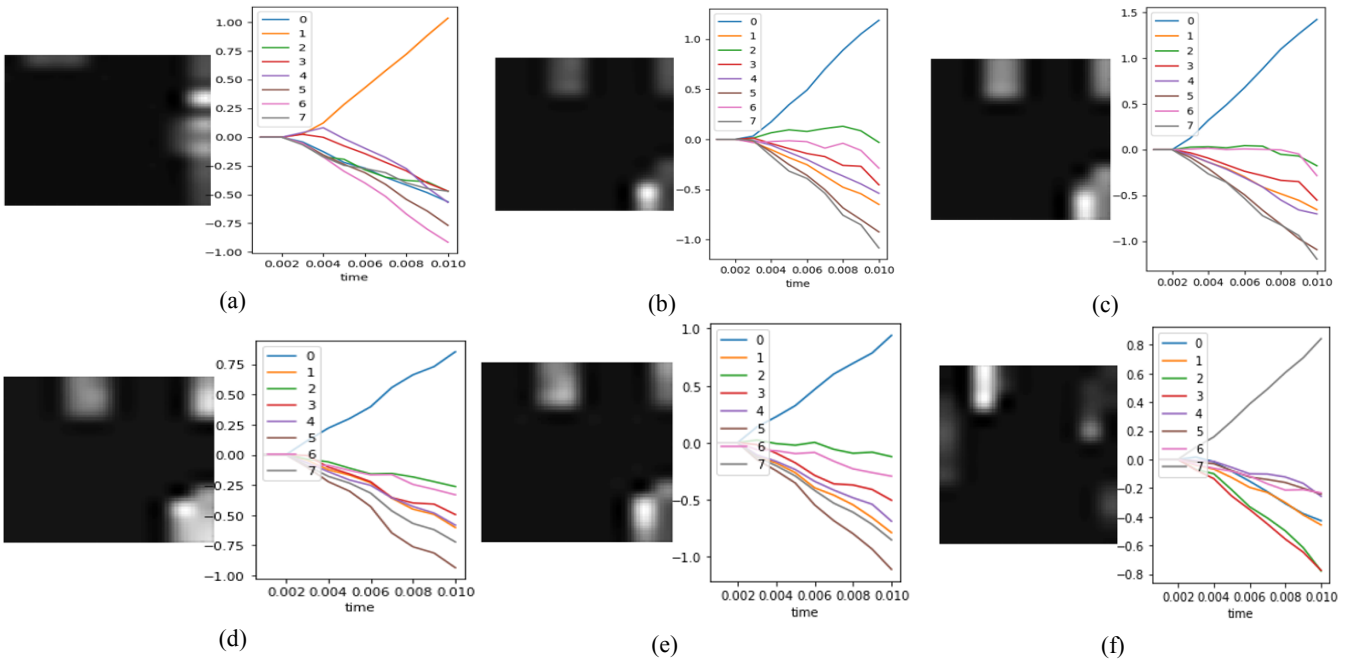


Fig. 4. CSNN classification results: (a), (b) and (c) generated using the Strathclyde gesture dataset (d), (e) and (f) generated using the CapgMyo dataset

higher prediction possibility begin to rise while other gesture lines going down. Eventually, only one line will be determined by the network to be the actual prediction result. According to figure 3, lines for gesture 1, 3 and 4 all rise up before 0.004s which indicates the network believes the input common energy-density map can represent these 3 gestures. However, after about 0.001s processing time, only gesture 1 line keeps rising and the other two begin to drop down which means the network prediction result for this input common energy-density map is gesture 1. It is also clearly shown in figure 4 and figure 5, that even for the same performed gesture (palm up), the common energy-density maps can appear to be different in some of the sensor locations. This is mainly because the muscle strength levels are unique for each subject. However, the network still makes the right prediction in 0.002 seconds. Our proposed network achieves 98.76% average hand gesture/movement recognition accuracy for the Strathclyde dataset.

B. CapgMyo Dataset

The CapgMyo open source hand gesture dataset used in the research aims for further evaluation of our approach.

The CapgMyo database contains three sub-databases (denoted as DB-a, DB-b and DB-c) in terms of the acquisition procedure. DB-a contains 8 isometric and isotonic hand gestures obtained from 18 of the 23 subjects. Each gesture in DB-a was held for 3 to 10 seconds. Every subject in DB-b contributed sEMG signals which the electrodes of the array were attached at slightly different positions each time. DB-c contains 12 basic movements of the fingers obtained from 10 of the 23 subjects. As some of the collection settings of the CapgMyo dataset are different compared with the Strathclyde dataset (length of the gestures' performing time, repeat performing numbers per gesture per subject, etc.), the signal

processing methodology applied is also different to guarantee the formed common energy-density maps share the same format. The signals were band-pass filtered at 20-380 Hz and sampled at 1,000 Hz. The resulting value was normalized to the $[-1, 1]$ range. Due to each of the sub-databases performs same gestures or hand movements, DB-a was selected for the testing process[13].

DB-a contained hand gestures are shown in Figure 3 and the testing results are presented in Figure 4(d), 4(e) and 4(f).

Like the Strathclyde dataset testing result figures, the numbers in each result figure here also represent a pre-defined hand gesture/movement (0: thumb up; 1: extension flexion; 2: flexion extension; 3: thumb opposing; 4: abduction; 5: fingers flexed; 6: pointing index; 7: adduction). The vertical axis of the left part of each figure still indicates the possibility of which gesture is the input common energy-density map represents and the horizontal axis refers to the processing time in seconds.

According to figure 4(a), 4(b), 4(c) the network begins to distinguish gestures at around 0.0025s. As mentioned in background part, different subjects' muscle strength levels would relate to different recorded signals' values. In this case, the recorded signals are weaker which indicates the features for input common energy-density maps are less evident compared with for common energy-density maps generated from the Strathclyde dataset. So longer processing time is required for the network to extra and learn the features. However, after the same 0.01s processing progress, the network produces the prediction.

For the CapgMyo dataset, the proposed convolutional spiking neural network achieves an average 98.21% hand gesture/movement recognition accuracy. Though the testing result is slightly lower compared to what was achieved with

the Strathclyde dataset, it still proves that the common energy-density map methodology and convolutional spiking neural network can be applied on various sEMG datasets.

C. Result comparison

The experimental results achieved by the convolutional spiking neural network were first compared with our previous results generated by IA-CNN methodology. Table III lists the comparison details.

TABLE III. Classification Results Comparison

Method	Gesture Number	Subject Number	Dataset	Result
IA-CNN	8	23	Strathclyde	98.96%
IA-CNN	8	10	CapgMyo	97.59%
CSNN	8	23	Strathclyde	98.76%
CSNN	8	10	CapgMyo	98.21%

According to Table III, for the used Strathclyde dataset, IA-CNN achieves 98.96% testing accuracy with 30 minutes training time while CSNN achieves only 0.2% less, 98.76% testing accuracy with 5 minutes training time. The small number of test losses is mainly due to the conversion process. As a pre-set threshold is required when transfer normal images into spike images, any lower value features would be ignored. In other words, the convolutional spiking neural network doesn't receive and learn from the information contained inside these features. However, as these kinds of information losses can't be avoided, the LIF model is applied in this research to minimize these losses.

After the comparison between our own work, we also compared the test results obtained by the convolutional spiking neural network with other researchers' previous achievements. The details are presented in Table IV with the number of gestures considered, and the number of amputees involved in the study.

TABLE IV. Classification Results Comparison for Multiple Methodologies

Method	Gesture Number	Subject Number	Input Data Format	Result
RCNN[14]	11	8 healthy people	linear spectral coefficients	90.6%
BPNN[22]	4	unknown	Mean Absolute Values	88.4%
ANN[23]	5	60 people	absolute value encoded through L-BFGS method	$85.08 \pm 15.21\%$
CNN[15]	8	unknown	HD-EMG muscle activation maps	98.15%
CSNN	8	9 healthy people 14 amputees	common energy-density map	98.76%
CNN[20]	6	unknown	wavelet transform	94%

As shown in Table IV, among all listed 6 different hand gesture recognition methodologies, the CSNN achieves the best 98.76% recognition accuracy. The second accurate

recognition with 98.15% is obtained by Simon Tam et al. through convolutional neural network[15]. The third high 94% hand gesture recognition accuracy is achieved by DC Oh et al. , also through convolutional neural network[14]. The fourth high accuracy is 90.6%, gained by Xiang Chen et al. through recurrent convolutional neural networks[16]. Both recognition results for BPNN (back propagation artificial neural network)methodology applied by Md. Rezwanul Ahsan et al. and artificial neural network designed by EA Chung et al. achieved above 85% accuracy [17, 18]. However, as the target gesture number and applied subject number are different (4 gestures collected from unknown number of subjects for BPNN methodology and 5 gestures collected from 60 people), it is difficult to evaluate which method obtains better recognition results.

IV. CONCLUSION

In this paper, we have presented a novel common energy-density map based method for personalizing EMG-based models with convolutional spiking neural network and we have evaluated it with our high-density sEMG dataset (Strathclyde dataset) and another open source sEMG dataset (CapgMyo dataset). The acquired gesture recognition performance is compared with some proposed prior work, which used different techniques such as wavelet transforms, Hudgins time Domain features and self-designed auto encoder with various machine learning and deep learning methodologies. Even though these experimental results were achieved using different categories and forms of different sEMG datasets, the collection instrument of all sEMG datasets remains the same i.e. the dataset used by others contain sEMG signals just like the sEMG dataset we have used. The core advantage of our approach lies in the minimisation of dataset processing time and network training time together with high hand gesture recognition accuracy. According to the literature, this advantage would become more significant if a spiking relu function was used instead of the LIF model[19]. Based on the Strathclyde dataset, a standard convolutional neural network requires 30 minutes to complete a training session while the proposed convolutional spiking neural network only uses 5 minutes. As shown by the hand gesture recognition results, our approach outperforms other approaches. The experimental results also indicate that the common energy-density map based convolutional spiking neural network approach grants faster user execution and reaction times than other approaches without recognition accuracy loss. Moreover, according to Table I and Table II, some of the sensors for distinct gestures are common which indicates the potential of recognizing different gestures by only using these common sensors. Thus, we will focus on identifying common sensors for all gestures to further minimise the data processing time and test the fully build system on hardware.

ACKNOWLEDGMENT

This work is supported by the University of Strathclyde. We would like to thank the assistance of all members in the deep

learning group, Centre for Signal and Image Processing (CeSIP), EEE department. The authors would also gratefully acknowledge the collectors of the CapgMyo dataset.

REFERENCES

- [1] A. V. W. Smith, A. I. Sutherland, A. Lemoine, and S. McGrath, "Hand gesture recognition system and method," ed: Google Patents, 2000.
- [2] K. Luan and T. Matsumaru, "Dynamic Hand Gesture Recognition for Robot Arm Teaching based on Improved LRCN Model," 2019: IEEE, pp. 1269-1274.
- [3] D. Farina *et al.*, "The extraction of neural information from the surface EMG for the control of upper-limb prostheses: emerging avenues and challenges," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 4, pp. 797-809, 2014.
- [4] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7420-7431, 2012.
- [5] R. Menon, G. Di Caterina, H. Lakany, L. Petropoulakis, B. A. Conway, and J. J. Soraghan, "Study on interaction between temporal and spatial information in classification of EMG signals for myoelectric prostheses," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1832-1842, 2017.
- [6] W. Ke, Y. Xing, G. Di Caterina, L. Petropoulakis, and J. Soraghan, "Intersected EMG heatmaps and deep learning based gesture recognition," 2020.
- [7] R. Jimenez-Fabian and O. Verlinden, "Review of control algorithms for robotic ankle systems in lower-limb orthoses, prostheses, and exoskeletons," *Medical engineering & physics*, vol. 34, no. 4, pp. 397-408, 2012.
- [8] P. McCool, "Surface myoelectric signal analysis and enhancement for improved prosthesis control," 2014.
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," 2012, pp. 1097-1105.
- [10] R. Vaila, J. Chiasson, and V. Saxena, "Deep convolutional spiking neural networks for image classification," *arXiv preprint arXiv:1903.12272*, 2019.
- [11] S. Dutta, V. Kumar, A. Shukla, N. R. Mohapatra, and U. Ganguly, "Leaky integrate and fire neuron by charge-discharge dynamics in floating-body MOSFET," *Scientific reports*, vol. 7, no. 1, pp. 1-7, 2017.
- [12] Y.-H. Liu and X.-J. Wang, "Spike-frequency adaptation of a generalized leaky integrate-and-fire model neuron," *Journal of computational neuroscience*, vol. 10, no. 1, pp. 25-45, 2001.
- [13] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Scientific reports*, vol. 6, p. 36571, 2016.
- [14] D. C. Oh and Y. U. Jo, "EMG-based hand gesture classification by scale average wavelet transform and CNN," International Conference on Control Robot System Society, pp. 533-538, 2019.
- [15] S. Tam, M. Boukadoum, A. Campeau-Lecours, and B. Gosselin, "A Fully Embedded Adaptive Real-Time Hand Gesture Classifier Leveraging HD-sEMG & Deep Learning," *IEEE transactions on biomedical circuits and systems*, 2019.
- [16] X. Chen, X. Zhang, Z.-Y. Zhao, J.-H. Yang, V. Lantz, and K.-Q. Wang, "Multiple hand gesture recognition based on surface EMG signal," 2007: IEEE, pp. 506-509.
- [17] M. R. Ahsan, M. I. Ibrahimy, and O. O. Khalifa, "Electromyography (EMG) signal based hand gesture recognition using artificial neural network (ANN)," 2011: IEEE, pp. 1-6.
- [18] E. A. Chung and M. E. Benalcázar, "Real-Time Hand Gesture Recognition Model Using Deep Learning Techniques and EMG Signals," 2019: IEEE, pp. 1-5.
- [19] M. Pfeiffer and T. Pfeil, "Deep learning with spiking neurons: opportunities and challenges," *Frontiers in neuroscience*, vol. 12, p. 774, 2018.